# NovoWorx User Manual

## Version 1.0

## EXOME WORKFLOW

# Table of Contents


1.0 Using the Multisample CSV file input....3
2.0 Starting your workflow....5
2.1 Creating your Project....5
Step 1....5
Step 2....6
Step 3....7
Step 4....8
2.2 Using the novoWorx job creation wizard....9
Step 1....9
Step 2....9
Step 3....10
Step 4....10
Step 5....11
Step 6....13
Step 7....14
3.0 Navigating a Running Project....15
4.0 The Job Status Panel....16
5.0 Viewing and Monitoring your results....16
6.0 Using the sidebar....17
7.0 Using the project job information view....18
8.0 Output Files....20
8.1 BAM....21
8.2 VCF....22
8.3 CSV & TSV & BED....23
8.4 HTML....24
8.5 TXT....25

## 1.0 Using the Multisample CSV file input

The exome multisample project workflow is designed to enable users to design and run a large set of data samples without having to input each sample individually. This is done using a comma separated value file which holds all the necessary information to set up the experiment

To create a Workflow, the csv file has to be populated with the following details.

sample_name,condition,format,type,file1,file2,mean,stddev

| Field Name | Description |
|---|---|
| sample_name | The individual sample name. This is a unique identifier for each sample. |
| condition | The sample condition . All samples with the same condition will be merged and will produce one BAM and one VCF per condition. |
| format | The fomat of the file. This is either fastq,csfastq, fasta or csfasta. |
| type | The type of read. This is either Mate-Pair, Pair-End, or Fragment. |
| file1 | The full path of the first read. |
| file2 (optional) | The full path of the second read. This may be left blank if only one read is present. |
| mean (optional) | The mean of the read length in base pairs. This field is optional and may be left empty |
| stddev (optional) | The Standard Deviation of the read length in base pairs. This field is optional and may be left empty. |

An example of a line in the CSV file is as below:-

Example of a sample with 2 reads:

```
NFLE,6417,fastq,paired-end,/home/user/6417_R1.fastq.gz,/home/user/6417_R2.fastq.gz,0,0
```

Example of a sample with only 1 read including the mean and standard deviation:

```
NFLE,6417,fastq,paired-end,/home/user/6417_R1.fastq.gz,,0,0
```

Example of a sample with only 1 read and NOT inclusive of the mean and standard deviation:

```
NFLE,6417,fastq,paired-end,/home/user/6417_R1.fastq.gz
```

Each line represents one sample in the experiment.

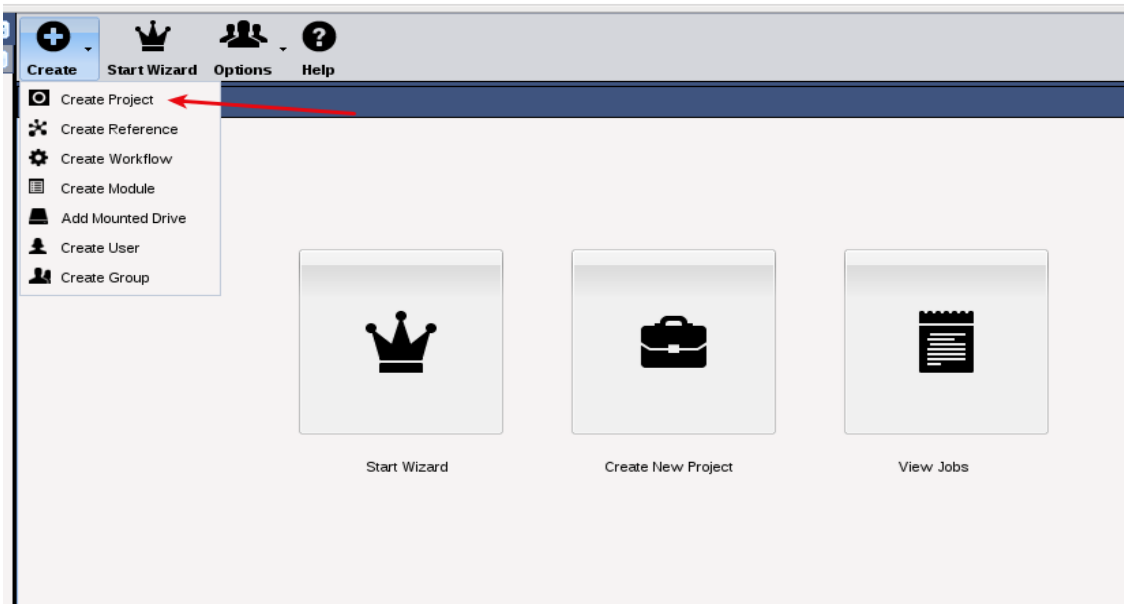
It should be noted that the novoWorx pipeline groups all the samples into the condition that they are placed in. If separate BAM files are necessary, they should be named uniquely as follows:

```
NFLE_6417,6417,fastq,paired-end,/home/user/6417_R1.fastq.gz,/home/user/6417_R2.fastq.gz,0,0
NFLE_6418,6418,fastq,paired-end,/home/user/6417_R1.fastq.gz,/home/user/6417_R2.fastq.gz,0,0
```

## 2.0 Starting your workflow

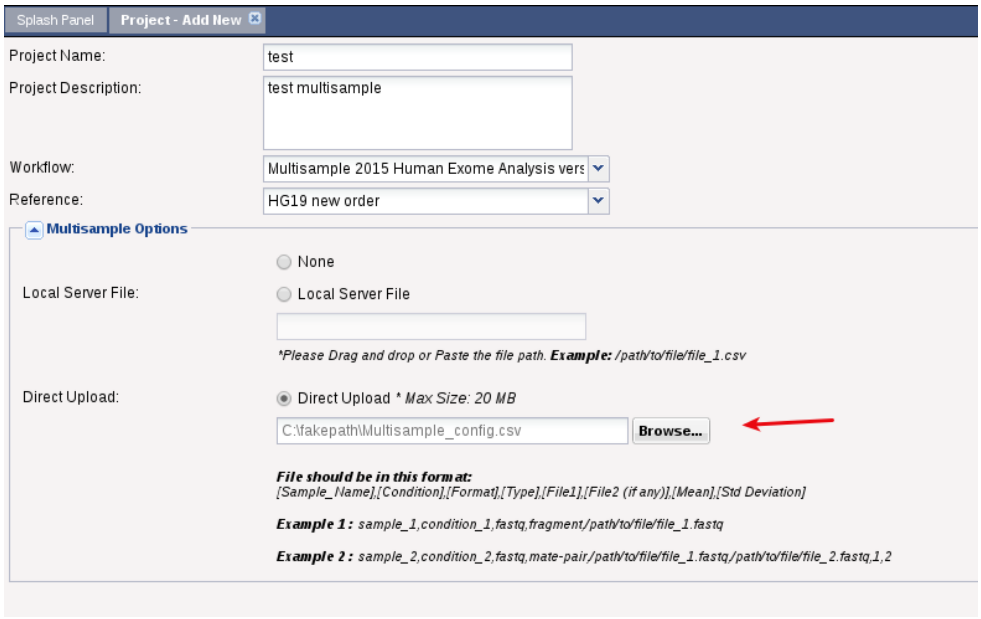
### 2.1 Creating your Project

#### Step 1




When logged on to novoWorx, click on the Create button and select Create Project

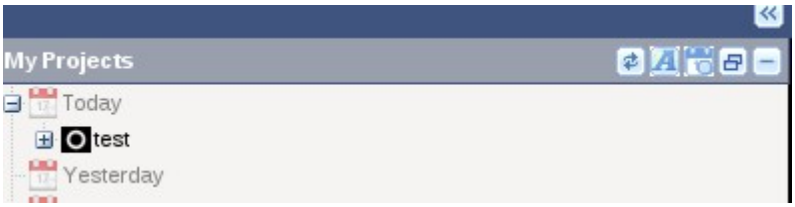
### Step 2




Fill in the details of the project as desired. Click on the Direct Upload radio button and browse for the CSV configuration file.

Once the CSV file is uploaded, click on the Save button at the bottom of the page. Click OK on the project creation confirmation dialog pop up.

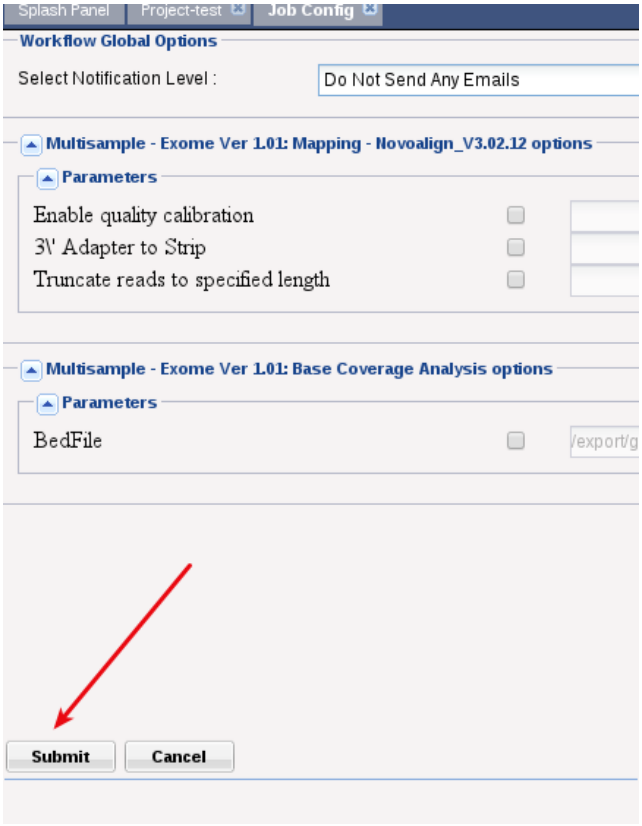
### Step 3




On the side bar, you will see the project that you have just created. Click on the name of the project and the splash panel will open a tab detailing the project information.

Click on Run Project  and you will proceed to the final step of the process.

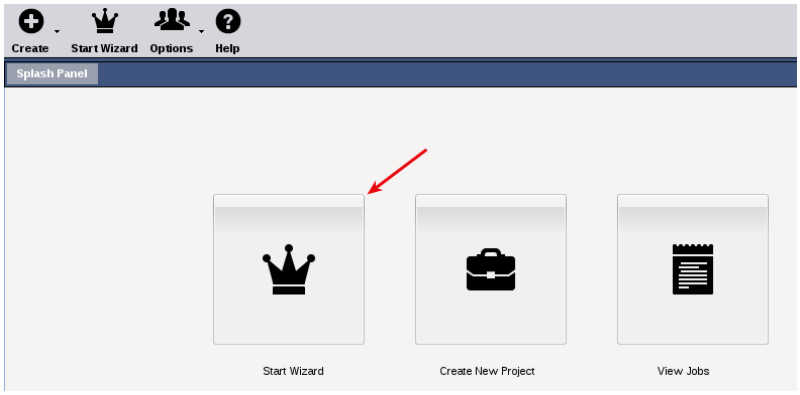
### Step 4




If you desire to use any additional options, you may do so in the final splash panel. Once all the desired options are selected, click the Submit button. A notification will pop up with the Job ID confirming that the job has been started.

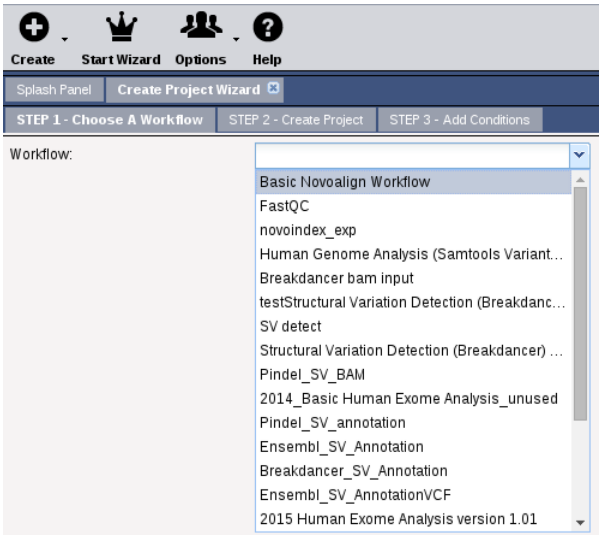
## 2.2 Using the novoWorx job creation wizard

### Step 1



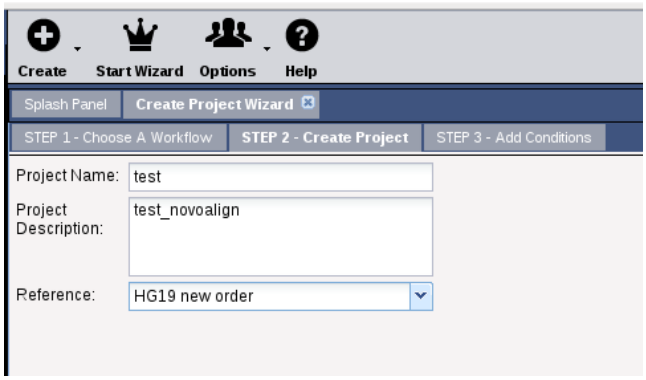

Click on the Start Wizard Button.

### Step 2




Select a Workflow from the drop down menu. Different workflows have different uses. Make sure you select the right one! Once this is done, click Next.

### Step 3



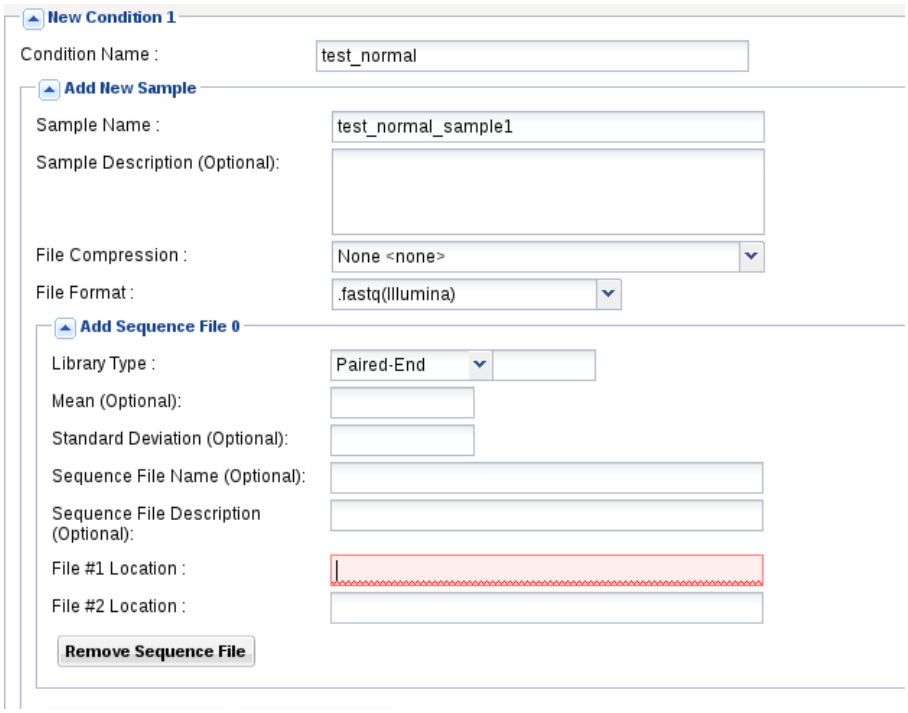

Here you can input the name and description for the project as well as select a suitable reference to use.  Click Next to continue.

### Step 4

Once the splash panel comes up, enter a condition name before clicking on Add a Sample. The condition name is not unique and samples with the same condition will be merged together.  This will mean that the samples with the same condition name will be merged into a single vcf file. If this is not desirable for your research, please separate the samples and provide unique names for each condition (for example : Tumor_1, Tumor_2, etc).

### Step 5

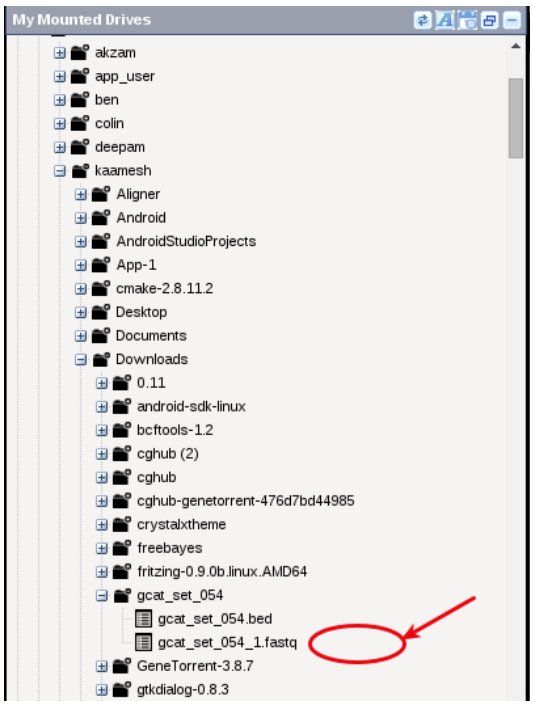
Once the sample is named, the File Compression must be selected. If no compression is used, select None. The file format is dependent on the sequencing technology in use at your facility. NovoWorx currently supports FASTA and FASTQ of both Illumina and ABI SOLiD Colorspace reads.




Select the Library type, whether single-ended or paired-end reads and the other optional reads may be left blank if no custom value is desired for the fields.

To input the read file location, one can key  in the location of the file or the file can be dragged and dropped from the mounted drives as shown in the following images:

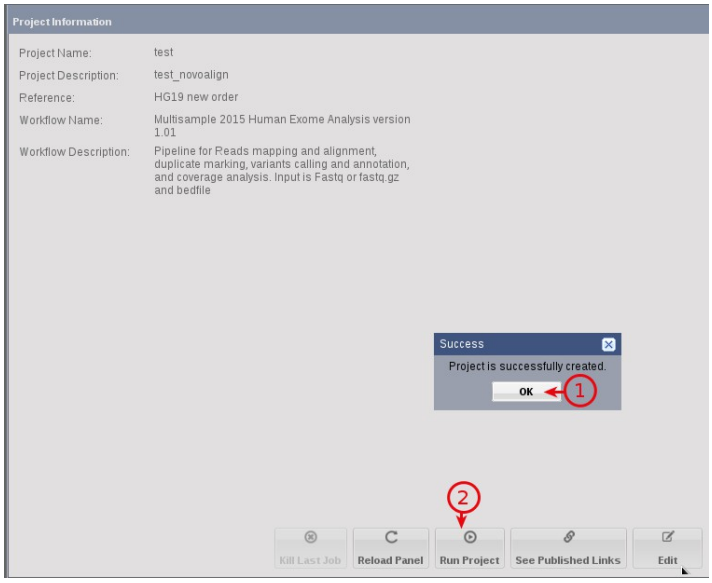
Click on My Mounted Drives to get an expandable list of drives and files mounted and navigate to the location of the desired file.




Click and drag the empty space next to the file name and drop it in the file location field. Do the same with the second file for paired-end data.

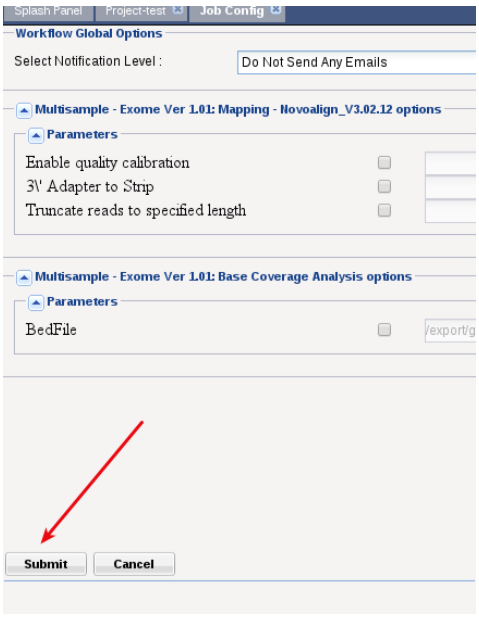
Click on Save Project to continue.

### Step 6

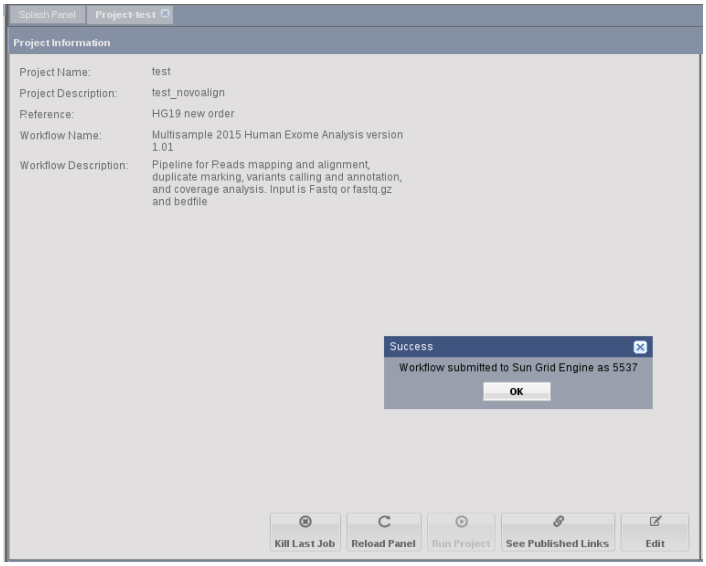



Once you click on Save Project, the job creation splash panel will close and the Project information splash panel will open with a dialogue box confirming the creation of the project. Click OK and then click on Run Project to start running your project.

### Step 7

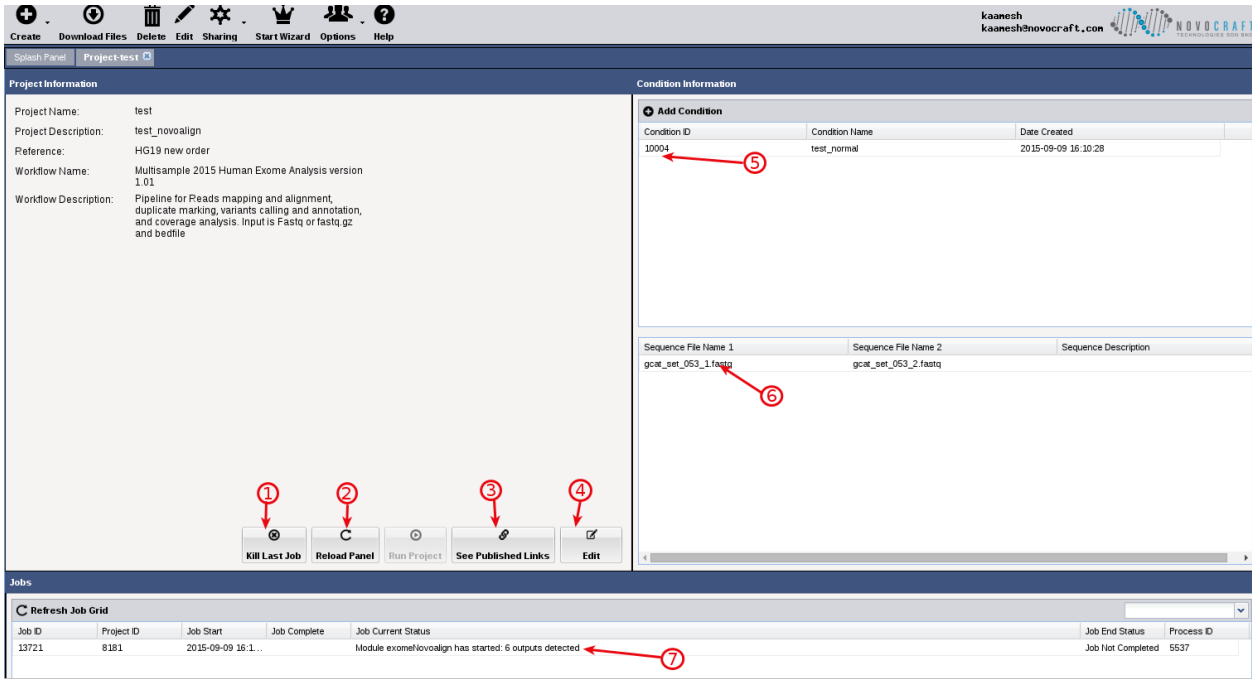



The final splash panel is where you input the final options for the project. If none of these options are desired or you would prefer the default values, leave them blank and click Submit.
You will get another dialog box letting you know that the job has been submitted and the job is now running.



## 3.0 Navigating a Running Project

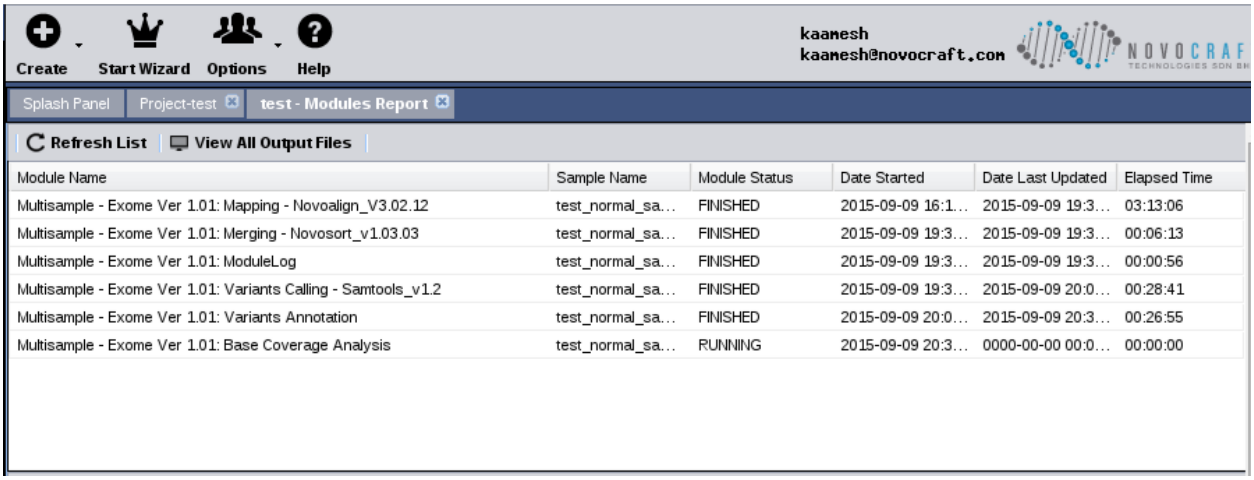
When logged on to novoWorx, clicking on the project name will bring up the project information splash panel. From here, multiple options such as viewing jobs, updating jobs or view output files are accessible as described below.




| Legend | Option | Description |
|---|---|---|
| 1 | Kill Last Job | Kill the most recent running job for the selected project. |
| 2 | Reload Panel | Reload the job status panel and update the running module as well as the job status. |
| 3 | Published Links | Open a splash panel that shows the list of downloadable files in the form of web links that can be downloaded via FTP. |
| 4 | Edit | Allows the user to edit the project name and description. |
| 5 | Condition Information | View details regarding the conditions already available and add conditions as the project is running. |
| 6 | Sequence Information | View the information regarding the sequence sets for each of your conditions. |
| 7 | Jobs | View information regarding the jobs that are currently running. |

## 4.0 The Job Status Panel

The Job Status Panel shows the list of jobs completed or currently running for a particular project. Double-clicking on the job will bring up a splash panel with the details of the modules that have been run to completion as well as the queued and running jobs.


Create
Start Wizard
Options
Help
kaanesh
kaanesh@novocraft.com
Splash Panel
Project-test
test - Modules Report
Refresh List
View All Output Files
Module Name | Sample Name | Module Status | Date Started | Date Last Updated | Elapsed Time
Multisample - Exome Ver 1.01: Mapping - Novoalign_V3.02.12 | test_normal_sa... | FINISHED | 2015-09-09 16:1... | 2015-09-09 19:3... | 03:13:06
Multisample - Exome Ver 1.01: Merging - Novosort_v1.03.03 | test_normal_sa... | FINISHED | 2015-09-09 19:3... | 2015-09-09 19:3... | 00:06:13
Multisample - Exome Ver 1.01: ModuleLog | test_normal_sa... | FINISHED | 2015-09-09 19:3... | 2015-09-09 19:3... | 00:00:56
Multisample - Exome Ver 1.01: Variants Calling - Samtools_v1.2 | test_normal_sa... | FINISHED | 2015-09-09 19:3... | 2015-09-09 20:0... | 00:28:41
Multisample - Exome Ver 1.01: Variants Annotation | test_normal_sa... | FINISHED | 2015-09-09 20:0... | 2015-09-09 20:3... | 00:26:55
Multisample - Exome Ver 1.01: Base Coverage Analysis | test_normal_sa... | RUNNING | 2015-09-09 20:3... | 0000-00-00 00:0... | 00:00:00

## 5.0 Viewing and Monitoring your results

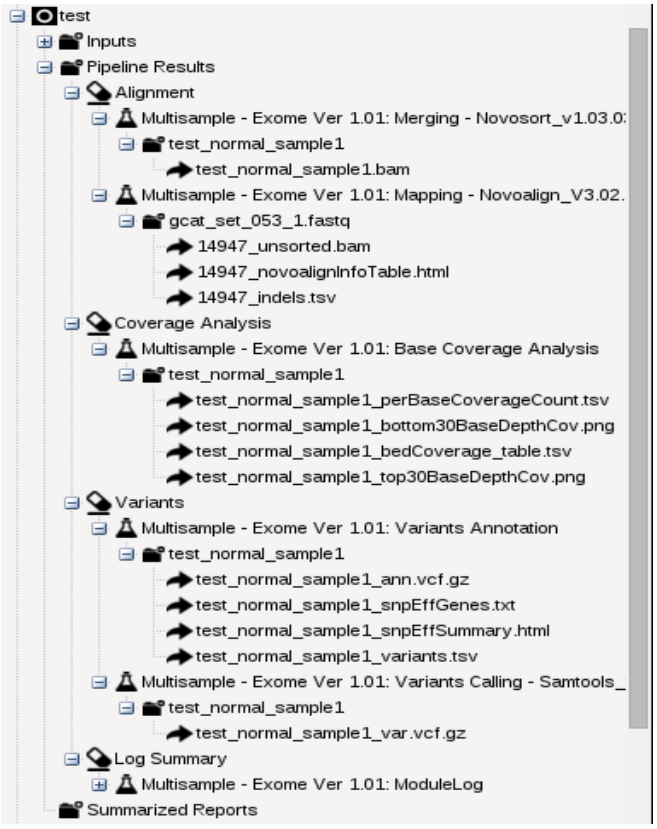
The results of the pipeline are grouped by the modules that have been run within the pipeline. Each module will have 3 sub-folders which are:- Inputs, Pipeline Results and Summarized Reports .

Each of these sub-folders has one sub-folder for each sample used in the module. Within each sample sub-folder is the output files for each of the modules which can be viewed in the novoWorx interface.

- The Inputs sub-folder has a list of the input files that have been used in the pipeline.
- The Pipeline Results sub-folder has the list of pipeline outputs for each module. A detailed Description of the type of files and the views for each file is detailed in the Output Files Section.
- The Summary Reports folder contains a summarized report for the entire pipeline that is compiled after conclusion. Not all of the pipelines have a summarized report and the folder may therefore be empty.

There are a few methods of viewing your results in the novoWorx user interface.

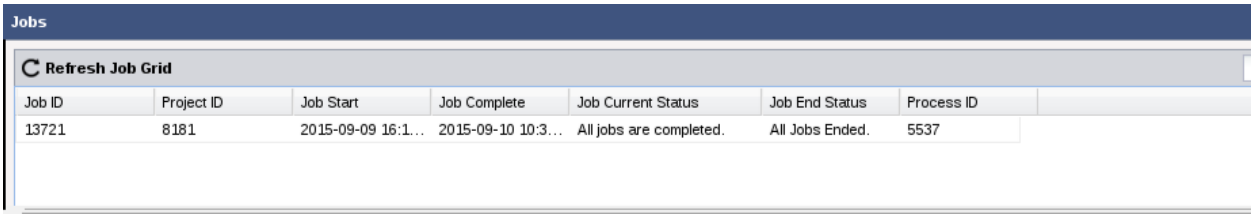
## 6.0 Using the sidebar




Expanding the side bar for each of the project folders reveals the results for each module in the pipeline. Clicking the file name will open a new tab in the main panel which will show the files in an easy to understand format.

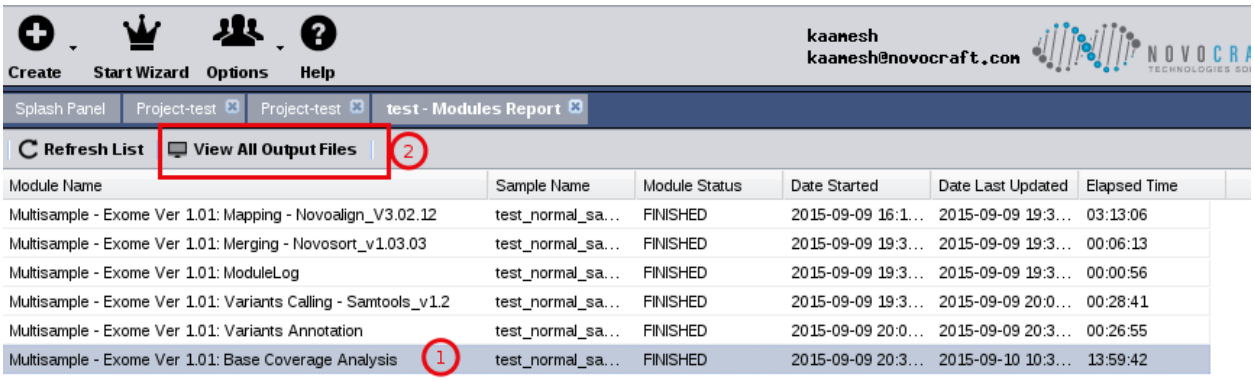
The sidebar is repopulated every time a module finishes. New outputs will be revealed as the modules are completed.

## 7.0 Using the project job information view



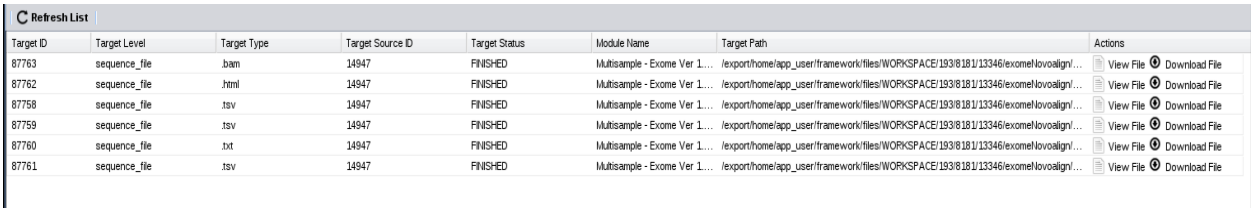

The project information job view shows a more comprehensive list of outputs for each job that has been run.

Double clicking on the Job ID that you wish to view will bring up a list of the completed modules or running modules as shown below:

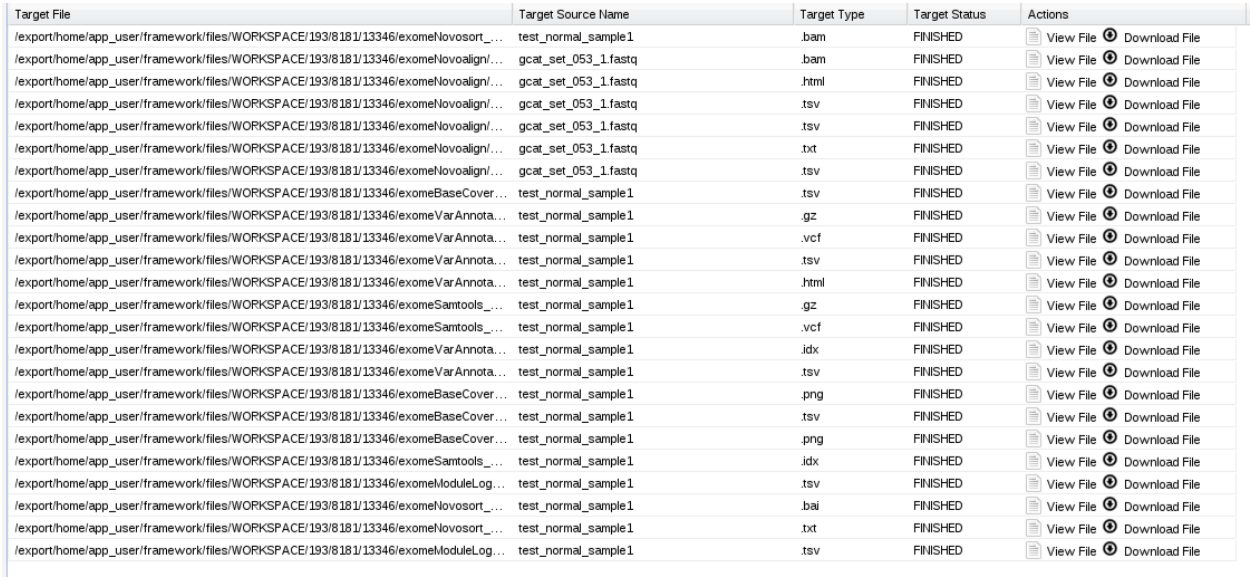



Based on the image above:-

1) Clicking on the results you would like to view will bring up a list of associated documents and/or files that can be viewed in novoWorx or downloaded for your reference as shown below:



2) You can also view all the files that have been produced by the pipeline by clicking on the View All Output Files button. This will open a new tab with the list of all the files not sorted by module as shown in the following image:-



| Target File | Target Source Name | Target Type | Target Status | Actions |
|---|---|---|---|---|
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovosort_... | test_normal_sample1 | .bam | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovoalign/... | gcat_set_053_1.fastq | .bam | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovoalign/... | gcat_set_053_1.fastq | .html | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovoalign/... | gcat_set_053_1.fastq | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovoalign/... | gcat_set_053_1.fastq | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovoalign/... | gcat_set_053_1.fastq | .txt | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovoalign/... | gcat_set_053_1.fastq | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeBaseCover... | test_normal_sample1 | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeVarAnnota... | test_normal_sample1 | .gz | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeVarAnnota... | test_normal_sample1 | .vcf | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeVarAnnota... | test_normal_sample1 | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeVarAnnota... | test_normal_sample1 | .html | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeSamtools_... | test_normal_sample1 | .gz | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeSamtools_... | test_normal_sample1 | .vcf | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeVarAnnota... | test_normal_sample1 | .idx | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeVarAnnota... | test_normal_sample1 | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeBaseCover... | test_normal_sample1 | .png | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeBaseCover... | test_normal_sample1 | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeBaseCover... | test_normal_sample1 | .png | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeSamtools_... | test_normal_sample1 | .idx | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeModuleLog... | test_normal_sample1 | .tsv | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovosort_... | test_normal_sample1 | .bai | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeNovosort_... | test_normal_sample1 | .txt | FINISHED | View File Download File |
| /export/home/app_user/framework/files/WORKSPACE/193/8181/13346/exomeModuleLog... | test_normal_sample1 | .tsv | FINISHED | View File Download File |

Clicking on View File will open the file in the novoWorx interface. Take note that not all files formats are supported by novoWorx. For a list of files that are supported, please navigate to the Output Files Section of this manual.

Files may also be downloaded for viewing outside of the novoWorx platform by clicking on the Download File button.

A description of the files available is included in the Output Files Section.

## 8.0 Output Files

NovoWorx is able to handle most standard bioinformatics file formats and is able to display them in an easy to process manner. Among the file formats supported by novoWorx are :-

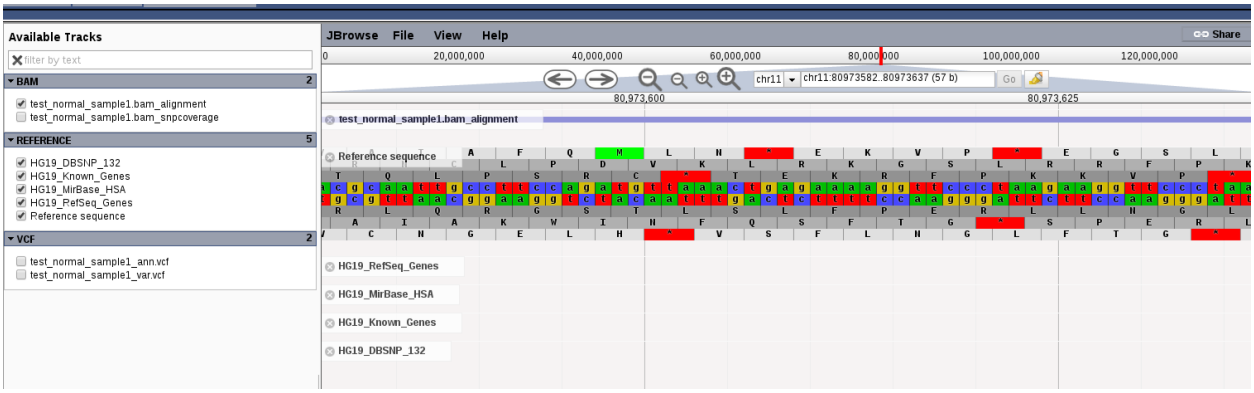
- BAM
- VCF
- CSV
- TSV
- HTML
- TXT
- BED

The following sections of this manual show examples of how the files are displayed in novoWorx.

### 8.1 BAM

A BAM file is a compressed file format that contains the alignments in a binary format. Further information regarding the BAM and SAM formatting can be obtained at the [SAMTOOLS format specification page.]

Double Clicking on the BAM file within the novoWorx interface will bring up the Jbrowse Genome Browser as shown in the image below:-

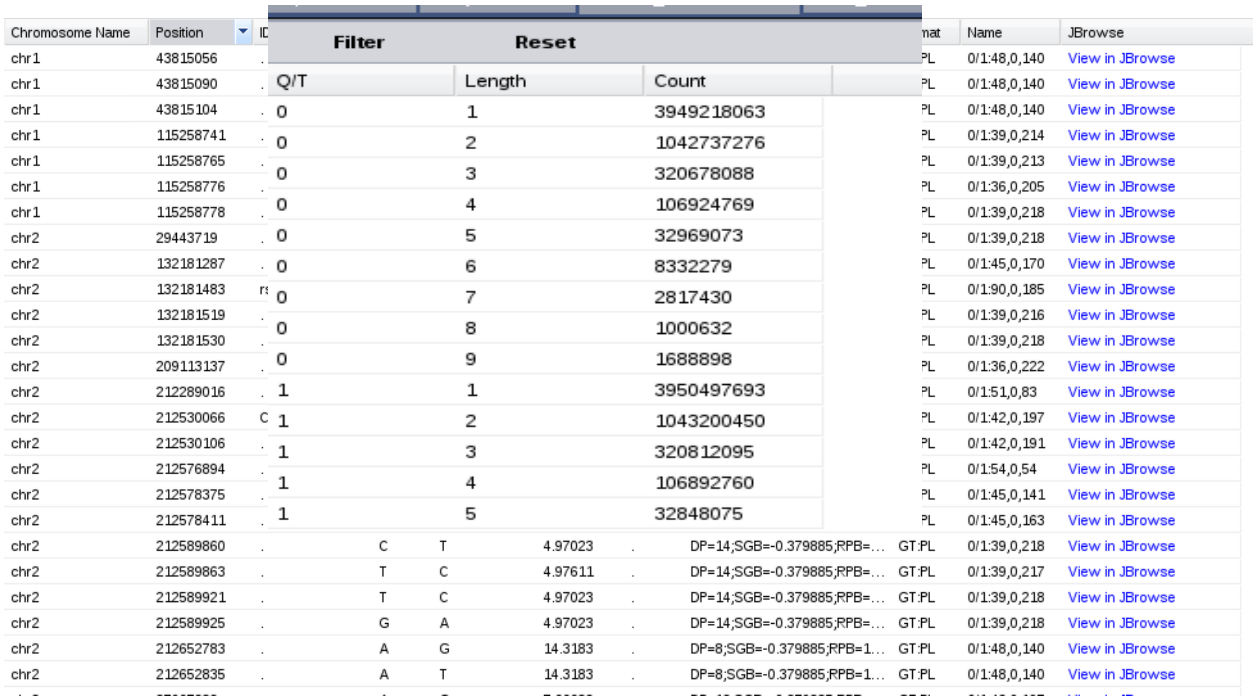



Here the BAM file can be viewed in tandem with the reference files. The BAM file is also downloadable for viewing in the Broad Institute Integrative Genomics Viewer (IGV). To download the BAM file, navigate to the Alignment section on the sidebar and click the sample file to view the aligned BAM file. Right click on the BAM file and select Download from the list of options.

You can also download the file by **navigating the Job Information panel.**

### 8.2 VCF

The VCF format is the standard format that variants are called and displayed in. For more information regarding the VCF format please visit the SAMTOOLS VCF 4.2 specification.



| Chromosome Name | Position | ID | | | | | | | Name | JBrowse |
|---|---|---|---|---|---|---|---|---|---|---|
| chr1 | 43815056 | . | | | | | | | 0/1:48,0,140 | View in JBrowse |
| chr1 | 43815090 | . | | | | | | | 0/1:48,0,140 | View in JBrowse |
| chr1 | 43815104 | . | | | | | | | 0/1:48,0,140 | View in JBrowse |
| chr1 | 115258741 | . | | | | | | | 0/1:39,0,214 | View in JBrowse |
| chr1 | 115258765 | . | | | | | | | 0/1:39,0,213 | View in JBrowse |
| chr1 | 115258776 | . | | | | | | | 0/1:36,0,205 | View in JBrowse |
| chr1 | 115258778 | . | | | | | | | 0/1:39,0,218 | View in JBrowse |
| chr2 | 29443719 | . | | | | | | | 0/1:39,0,218 | View in JBrowse |
| chr2 | 132181287 | . | | | | | | | 0/1:45,0,170 | View in JBrowse |
| chr2 | 132181483 | rs | | | | | | | 0/1:90,0,185 | View in JBrowse |
| chr2 | 132181519 | . | | | | | | | 0/1:39,0,216 | View in JBrowse |
| chr2 | 132181530 | . | | | | | | | 0/1:39,0,218 | View in JBrowse |
| chr2 | 209113137 | . | | | | | | | 0/1:36,0,222 | View in JBrowse |
| chr2 | 212289016 | . | | | | | | | 0/1:51,0,83 | View in JBrowse |
| chr2 | 212530066 | C | | | | | | | 0/1:42,0,197 | View in JBrowse |
| chr2 | 212530106 | . | | | | | | | 0/1:42,0,191 | View in JBrowse |
| chr2 | 212576894 | . | | | | | | | 0/1:54,0,54 | View in JBrowse |
| chr2 | 212578375 | . | | | | | | | 0/1:45,0,141 | View in JBrowse |
| chr2 | 212578411 | . | | | | | | | 0/1:45,0,163 | View in JBrowse |
| chr2 | 212589860 | . | C | T | 4.97023 | . | DP=14;SGB=-0.379885;RPB=… | GT:PL | 0/1:39,0,218 | View in JBrowse |
| chr2 | 212589863 | . | T | C | 4.97611 | . | DP=14;SGB=-0.379885;RPB=… | GT:PL | 0/1:39,0,217 | View in JBrowse |
| chr2 | 212589921 | . | T | C | 4.97023 | . | DP=14;SGB=-0.379885;RPB=… | GT:PL | 0/1:39,0,218 | View in JBrowse |
| chr2 | 212589925 | . | G | A | 4.97023 | . | DP=14;SGB=-0.379885;RPB=… | GT:PL | 0/1:39,0,218 | View in JBrowse |
| chr2 | 212652783 | . | A | G | 14.3183 | . | DP=8;SGB=-0.379885;RPB=1… | GT:PL | 0/1:48,0,140 | View in JBrowse |
| chr2 | 212652835 | . | A | T | 14.3183 | . | DP=8;SGB=-0.379885;RPB=1… | GT:PL | 0/1:48,0,140 | View in JBrowse |

Filter Reset

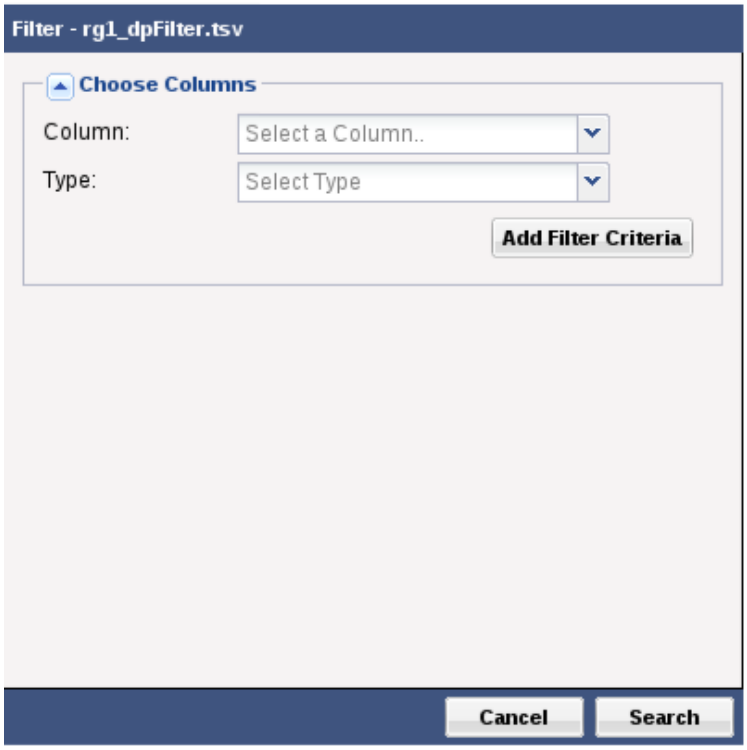
| Q/T | Length | Count |
|---|---|---|
| 0 | 1 | 3949218063 |
| 0 | 2 | 1042737276 |
| 0 | 3 | 320678088 |
| 0 | 4 | 106924769 |
| 0 | 5 | 32969073 |
| 0 | 6 | 8332279 |
| 0 | 7 | 2817430 |
| 0 | 8 | 1000632 |
| 0 | 9 | 1688898 |
| 1 | 1 | 3950497693 |
| 1 | 2 | 1043200450 |
| 1 | 3 | 320812095 |
| 1 | 4 | 106892760 |
| 1 | 5 | 32848075 |

In novoWorx, the VCF format is made easy to read by splitting the tabs into columns that can be sorted. The VCF file is also viewable in Jbrowse to show the location of the variant in the aligned BAM file. Clicking on the button opens the Jbrowse tab and jumps to the location of the variant.

### 8.3 CSV & TSV & BED

A CSV file is a file where the separator between fields is a comma and a TSV file is a file where the separator between fields is a tab. A BED file is a tab separated file that is generally comprised of a list of genomic range.

NovoWorx separates each field into an easy to view format. NovoWorx also makes searching for values easier using the Filter function by specifying the fields as shown below:



### 8.4 HTML

As novoWorx is a browse-based interface, HTML files will be opened in a new Browser tab and can be viewed in the native HTML format.

# Novoalign Report

| MODULE | ATTRIBUTES | 14947 |
|---|---|---|
| **Novoalign** | Version: | V3.02.12 - Build Feb 13 2015 @ 12:41:35 |
| | Licenced To: | Novocraft Technologies Sdn Bhd (Internal Use Only) |
| | Input Type: | Paired-End |
| | Input Format: | fastq |
| | Input Files Interpretation: | Illumina FASTQ, Casava Pipeline 1.8 & later. |
| | Input Files Name: | gcat_set_053_1.fastq<br>gcat_set_053_2.fastq |
| | Commands: | --hlimit 8<br>-p 5,20<br>--softclip 20<br>-d HG19_new_order.idx<br>-K 14947_mismatch.tsv<br>-o SAM<br>--hpstats 14947_indels.tsv<br>-H 20<br>-c 16<br>-t 20,4<br>-f gcat_set_053_1.fastq gcat_set_053_2.fastq |
| | Index Build Version: | 3.0 |
| | Hash Length: | 14 |
| | Step Size: | 2 |
| | Paired Reads: | 45038905 |
| | Proper Pairs: | 44430521 (98.6%) |
| | Read Sequences: | 90077810 |
| | Unique Alignment: | 79289074 (88.0%) |
| | Multi Mapped: | 10035642 (11.1%) |

### 8.5 TXT

The TXT format is a basic non-formatted text file. In novoWorx, the txt. files are displayed in a standard tab with no formatting and will look like you opened the file with any standard text editor. NovoWorx does NOT allow editing for text files.